Centre de Palethnologie Stratigraphique "ERURI"

DIALEKTIKÊ

CAHIERS DE TYPOLOGIE ANALYTIQUE

1975

Centre de Palethnologie
Stratigraphique «ERURI»
Maison d'Ossau
64260 ARUDY

M. Gallet

G. Laplace

M. Livache

R. Vilain

Publié avec le concours de l'Université de PAU
- Institut Universitaire de Recherche Scientifique -

ERRATA CORRIGE

p. 8 - ligne 19 : $\widehat{d}$'individus

" 9 - " 4 : $X^2 = \dfrac{(a - np)^2}{np} + \dfrac{(b - \widehat{nq})^2}{nq}$

" 9 - " 19 : $X^2 = \dfrac{(a - \frac{\widehat{a+b}}{2})^2}{n \frac{(a+b)}{2n} \frac{(c+d)}{2n}}$

" 15 - " 23 : $\underline{nq} = n(\widehat{1-p})$

" 16 - " 16 : $X^2 = \widehat{\sum} \dfrac{O^2}{C} - \widehat{N}$ ou $X^2 = \widehat{\sum} \dfrac{(O-C)^2}{C}$

" 17 - " 20 :

	J_1	J_2	
I_1	a	b	a+b
I_2	c	d	c+d
	a+c	$\widehat{b+d}$	N

" 18 - " 19 : $P(O) = \dfrac{530!\ 383!\ 150!\ 3!\ \widehat{\ }}{533!\ 383!\ 147!\ 3!\ 0!}$

" 28 - " 12 : $d(X_h, X_i)$ SUP $d(\widehat{X_h}, X_e), d(X_e, X_i)$"

" 28 - " 24 : figurant les $\widehat{\text{feuilles}}$

Sommaire

GALLET M. – Premiers éléments pour une étude des techniques de débitage..... p. 1

LAPLACE G. – Précisions sur la démarche de l'analyse structurale..... " 8
LIVACHE M.

LAPLACE G. – Distance du Khi 2 et algorithmes de classification hiérarchique..... " 22

VILAIN R. – La fiche de typologie analytique.... " 38

et quelques dessins humoristiques de J. BARANDIARAN

Responsable :
Delia Brusadin-Laplace
"Le Saule" – Rue de Tisserands
64800 COARRAZE

PREMIERS ELEMENTS POUR UNE ETUDE DES TECHNIQUES DE DEBITAGE

Michel Gallet

Tout travail de transformation des matériaux obéit à des lois : la technique moderne connaît, par exemple des lois de la coupe industrielle des métaux sur machines-outils. Ces lois font intervenir :

1) la nature et les caractéristiques mécaniques du matériau travaillé,

2) les processus techniques employés et toutes les conditions particulières d'application de ces techniques. De plus, il est possible de retrouver dans l'analyse du produit fini les modes particuliers de son élaboration.

C'est aussi le cas pour les "outils" préhistoriques dont une analyse fine permet parfois de retrouver les étapes et les conditions de fabrication. C'est certainement le cas pour le débitage qui est le premier stade de la fabrication de l'outil : le produit débité acquiert des caractéristiques différentes suivant le mode de débitage employé et il est très probablement possible à partir de l'étude de ces caractéristiques, de distinguer de façon systématique (et non plus seulement intuitive) les techniques différentes.

Il faut préciser que l'on doit entendre dans le mode de débitage, non seulement le choix du mode de percussion (direct ou indirect, par pression, au percuter dur ou tendre...) mais l'ensemble du processus de fabrication, c'est-à-dire : la préparation du nucleus et sa forme, la préparation du ou des plans de frappe, la position des points d'impact par rapport aux arêtes qui marquent le nucléus, la position successive de ces points d'impact sur le ou les plans de frappe, etc...

Quelques remarques.

1 - Une telle étude est rendue difficile parce que :

1) peu d'ensembles industriels nous livrent la totalité des produits débités à partir d'un même nucléus,

2) dans un même ensemble, plusieurs techniques peuvent avoir été employées en vue d'obtenir des produits différents,

3) bon nombre des produits débités ont été transformés en outils qui ne permettent plus de retrouver toutes les caractéristiques primitives de l'éclat. Dans cette dernière perspective, il faut même ajouter que les produits manquants sont logiquement ceux qui répondaient davantage au "schema" recherché par l'artisan.

Malgré ces difficultés, l'étude des produits de débitage paraît possible. La première démarche consiste à dégager les paramètres vraiment significatifs qui doivent être pris en compte.

Les premiers travaux devront porter à la fois et comparativement sur des séries, les plus complètes et homogènes possibles, provenant d'ensembles industriels stratigraphiquement bien définis, et sur des séries "modernes" débitées en laboratoire dans des conditions parfaitement définies et dont les produits pourront être recueillis en totalité et dans leur ordre chronologique d'obtention.

2 - Une autre difficulté est celle de l'abord métrologique de produits dont les formes sont toujours irrégulières pour ne pas dire "fantaisistes". Mesurer la largeur d'une lame peut consister à prendre une largeur au milieu, ou bien la largeur maxi, ou bien multiplier les mesures à l'infini pour en prendre la moyenne. Aucune solution n'est satisfaisante. La mesure ici doit être un choix délibéré mais bien précis de façon à permettre des comparaisons entre travaux de chercheurs différents. Les choix que propose cette petite étude, font suite à une série d'essais. Ce sont des suggestions, mais ils tiennent compte de certains impératifs sans lesquels la mesure n'est qu'une illusion de précision. Ils tentent en particulier d'organiser les mesures autour d'axes et de références précis.

3 - Les études entreprises portent exclusivement (pour l'instant) sur les produits débités à partir de nucléus du paléo-sup.(prismatiques, pyramidaux,...). Après plusieurs essais il est apparu important de pouvoir étudier séparément et donc de distinguer les produits issus de ces débitages.

Il est vrai que l'on trouve, transformés en outils, les supports les plus divers : des éclats de décorticage, des lames à crêtes aussi bien que des lames plus régulières façonnées "en série" sur les traces en cannelures d'enlèvements antérieurs longitudinaux. Mais outre que les caractéristiques mesurées sur l'ensemble de ces dernières sont différentes de celles des lames à crête ou comportant une partie de crête ou de cortex, l'étude de l'outillage du gisement étudié (1) a permis de mettre en évidence le choix de certains supports en vue de la fabbrication d'outils spécifiques.

exemple : 90% des LD, PD, DT, comportent au moins 2 cannelures sur la face dorsale, alors que le pourcentage des autres types fabriqués sur ce même support n'atteint pas 50%. Pour les grattoirs, ce pourcentage est de 35%.

Si l'on conçoit tout le processus préparatoire de façonnage d'un nucléus (du paléo-sup.) comme visant à obtenir finalement des "lames en série", il faut définir ce produit final sur des critères techniques (excluant délibérément le rapport longueur/largeur, et donc incluant éventuellement des produits courts comme des lames dites réfléchies). Nous avons défini ainsi la lame : produit débité après les enlèvements de préparation et comportant au minimum une arête longitudinale entre deux cannelures d'enlèvements préalables selon l'axe de la pièce (enlèvements totaux ou partiels).

Cette définition paraîtra très restrictive. Elle l'est. Notre but n'est pas d'exclure la possibilité d'une étude globale de l'ensemble des produits. Mais nous voyons ici l'avantage de pouvoir étudier séparément les produits les plus élaborés d'une technique, et de permettre ainsi des comparaisons plus fines avec les produits issus d'une autre industrie , et peut-être d'une autre technique.

4 - Une étude systématique du débitage devrait être complétée par une étude des caractéristiques des matériaux mis en oeuvre. Il faudrait pouvoir réaliser, pour chacun, les essais classiques de dureté, résistance à la compression, au choc, à la flexion, à la torsion, etc...

On peut toujours rêver des crédits nécessaires à l'installation de laboratoires comportant des équipements dans ce but !

---

(1) - La Rouvière (Ardèche) - B.S.P.F., tome 68, 1971.Etudes et travaux - fasc. 1.
- Les pièces à cran de la Rouvière - Etudes préhistoriques, n.5 juin 1973.

Les paramètres proposés.

1 - Longueur / L : mesurée selon l'axe du produit de débitage (mesure à 0,5 mm près).

2 - Largeur moyenne / l : la pièce est posée sur le graphique n.1 en fonction de sa Longueur.
L'axe de la pièce est divisée en 4 parties égales. On mesure alors au pied à coulisse la largeur proximale, la largeur médiane, la largeur distale et éventuellement la largeur maxi si celle-ci est différente des 3 premières (mesure à 0,2 mm près). La largeur moyenne est la moyenne des 3 ou 4 mesures.

3 - Epaisseur moyenne / e : aux mêmes repères que précédemment, on mesure les épaisseurs proximale, médiane, distale et éventuellement maxi.

4 - Parallélisme de direction des profils latéraux : on appelle direction d'un profil la droite donnant l'orientation d'un bord de la pièce.

4.1 - Par convention, c'est, dans le cas général, la droite passant par les points de ce bord situés au quart et au trois quarts de la Longueur. La pente des deux profils latéraux est donnée en pourcentage par la relation $100 \frac{ld - lp}{0,5\ L}$
Cette pente est donc marquée du signe + si les profils sont divergents, et du signe - si les profils sont convergents. On procède de cette façon même si la largeur maxi est située entre les largeurs distale et proximale (Fig. 2). Le bombé éventuel d'un bord sera pris en compte avec le paramètre n.5.

4.2 - Si le bord de la pièce venait à couper la droite, définie précédemment, sur l'un des quarts extrêmes proximal ou distal, un graphique d'angles, dont les pentes s'echelonnent par accroissement de 5%, donne directement la valeur de la pente de la pièce (Fig.3).

4.3 - Les lames sont ainsi séparées en classes : de 0 à 5%, de 5 à 10%, de 10 à 15% etc... et affectées d'un signe.

5 - Indices de rectitude des profils : un indice concerne le bord dextre, un autre le bord senestre. Il me paraît suffisant, pour l'instant, de placer la pièce sur un papier millimétré et d'évaluer à l'oeil (à 0,5 mm près) l'écart maximal des profils réels

de la pièce par rapport à la droite utilisée dans le paragraphe précédent, comme direction du profil (Fig. 4). L'indice est alors calculé en rapportant l'écart trouvé à la longueur servant de référence. Le résultat est donc donné en pourcentage. Il est positif pour un bord convexe, négatif pour un bord concave. Enfin il comporte une valeur positive et une valeur négative pour les bords sinueux.

6 - Types de talons : classés selon la typologie habituelle. Il faudra cependant préciser davantage, dans l'avenir, ce qui concerne la préparation technique du point d'impact et ajouter la mesure d'angle.

7 - Morphologie des extrémités proximales et distales : ceci reste à préciser.

8 - Morphologie de la face dorsale : cette face s'organise autour de une ou plusieurs arêtes principales. On peut retenir plusieurs types :

une arête = a
deux arêtes = b
trois arêtes = c
deux arêtes très proches = d (ce type tend vers a)

Lorsque l'organisation est différente depuis l'extrémité proximale en allant vers l'extrémité distale, on pourra noter, sous la forme a→b (cas d'une arête principale se dédoublant - voir figure 1).

9 - Sens des enlèvements précédents sur la face dorsale : on note :

- d'une part, les pièces pour lesquelles l'ensemble de ces enlèvements a été pratiqué dans le sens du débitage de l'éclat,
- d'autre part, les pièces dont au moins un enlèvement même partiel est en sens inverse,
- enfin, les pièces pour lesquelles l'examen ne permet pas d'être affirmatif.

Remarque : Cette notation paraît intéressante, on peut en effet se poser la question d'une relation entre la forme de cambrure du revers de la pièce et la présence d'enlèvements de sens opposés. Cette hypothèse est suggérée par l'étude du gisement de La Rouvière, où certains types d'outils, les pièces à cran par exemple, semblent avoir été fabriqués de façon préférentielle sur des supports présentant des traces d'enlèvements opposés.

10 - Cambrure de la face plane : cette cambrure met en cause l'équilibre dynamique des pièces susceptibles de faire partie d'un outil de jet par exemple. Un tel outil n'a pu être fonctionnel sans une cambrure extrêmement réduite ou un profil sinueux rétablissant

l'équilibre des masses. Certains outils par contre (des grattoirs par exemple) présentent parfois un profil concave de la face plane.

Les références de mesures sont choisies sur la face plane au quart ou au trois quarts de la Longueur (Fig. 5). Un petit appareil mis au point pour cette mesure, permet à la fois de mettre la pièce en place sur les deux appuis en fonction de sa Longueur, et de prendre trois mesures : dans les zones proximale, médiane et distale. Les mesures donnent la valeur des écarts des trois points mesurés par rapport au plan passant par le sommet des appuis. Ces valeurs peuvent être positives, négatives ou nulles. On en déduit le profil de la face plane. Pour comparaison, les valeurs d'écart sont rapportées à la Longueur de la pièce.

o o o o

Cette étude n'est que le point de départ d'une réflexion qui doit permettre de préciser et améliorer les méthodes, les paramètres choisis, et les moyens de mesure. Les critiques pourront certainement y aider, et elles sont souhaitées. Toutefois, les critères retenus après plusieurs séries de mesures peuvent permettre des premières comparaisons et ce sont les résultats de celles-ci qui nous permettront de juger de la valeur des premiers.

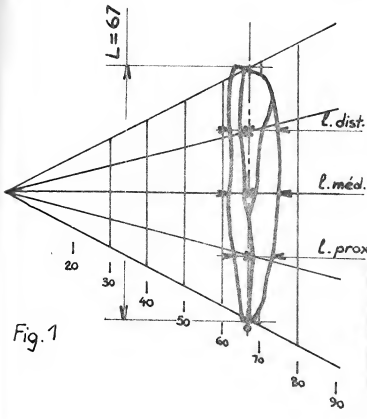


Fig. 1

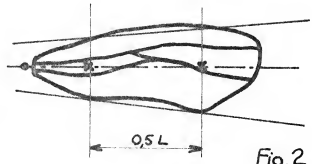


Fig. 2

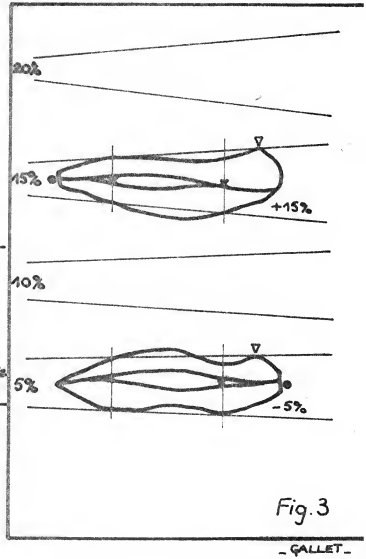


Fig. 3

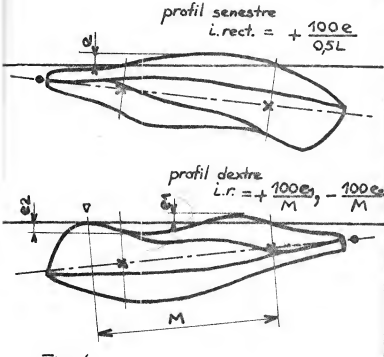


Fig. 4

- GALLET -

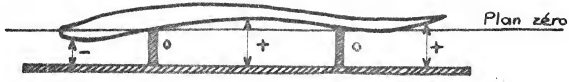


Fig. 5

PRECISIONS SUR LA DEMARCHE DE L'ANALYSE STRUCTURALE

Georges Laplace et Michel Livache

Depuis la publication de "De la dynamique de l'analyse structurale ou la typologie analytique",[+] de nouvelles recherches ont conduit à modifier la démarche de l'analyse structurale. Ces perfectionnements ont porté sur :

a - la définition d'un critère de contingence permettant de fixer les discontinuités ou ruptures dans la séquence structurale;
b - l'application de l'estimation de fréquences théoriques à partir des fréquences observées dans la séquence structurale ;
c - l'analyse structurale comparée ;
d - l'analyse de la dynamique structurale.

A. CRITERE DE CONTINGENCE ET SEQUENCE STRUCTURALE

1. Formules du critère de contingence.

1.1. Le critère de contingence quadratique (np et $nq \geqslant 10$).

Si nous considérons un échantillon de n individus où a individus possèdent un caractère donné et où $b = n-a$ ne l'ont pas, nous nous proposons de savoir si cette observation est compatible avec une certaine proportion théorique p, donnée à priori d'individus présentant le caractère donné dans la population.

Pour celà, on calcule :

[+] LAPLACE G. - De la dynamique de l'analyse structurale ou la typologie analytique. Rivista di Scienze Preistoriche, XXIX, 1974 p.1-71.

a - les effectifs théoriques de chacune des classes :
np et n(1-p) = nq

b - la quantité $\sum \frac{(O - C)^2}{C}$, c'est-à-dire :

$$X^2 = \frac{(a - np)^2}{np} + \frac{(b - nq)^2}{nq}$$

qui s'exprime plus simplement par la formule :

$$X^2 = \frac{(a - np)^2}{npq}$$

permettant de comparer une fréquence observée à une fréquence théorique par la méthode du Khi 2.

Soient a et b les effectifs de deux catégories de la séquence structurale, n la somme des effectifs de toutes les catégories, c et d les effectifs complémentaires de a et b tels que :

$c = n - a$ et $d = n - b$ d'où le tableau de contingence:

a	b	a+b
c	d	c+d
n	n	2n

Calculons les effectifs théoriques de chacune des classes :

$$np = \frac{n(a+b)}{2n} = \frac{a+b}{2} \quad \text{et} \quad nq = \frac{n(c+d)}{2n} = \frac{c+d}{2}$$

La formule $X^2 = \frac{(a-np)^2}{npq}$ devient :

$$X^2 = \frac{(a - \frac{a+b}{2})^2}{n \frac{(a+b)}{2n} \frac{(c+d)}{2n}} = \frac{(a - \frac{a+b}{2})^2}{\frac{n(a+b)(c+d)}{4n^2}} = \frac{4n \left(\frac{a-a+b}{2}\right)^2}{(a+b)(c+d)}$$

$$= \frac{4n \left(\frac{2a - a - b}{2}\right)^2}{(a+b)(c+d)} = \frac{4n \left(\frac{a - b}{2}\right)^2}{(a+b)(c+d)} \quad \text{d'où}$$

la formule du critère de contingence quadratique établie par G.Laplace:

$$X^2 = \frac{n(a-b)^2}{(a+b)(c+d)} \qquad (A.1.1.1)$$

et la formule du coefficient de corrélation de points :

$$\phi = \sqrt{\frac{\chi^2}{n}} \quad \text{soit} \quad \phi = \sqrt{\frac{(a-b)^2}{(a+b)(c+d)}} \qquad (A.1.1.2)$$

La formule du critère de contingence quadratique ne s'applique que lorsque les effectifs théoriques $\frac{a+b}{2}$ et $\frac{c+d}{2}$ sont supérieurs ou égaux à 10.

1.2. Utilisation de la loi binomiale (np ou nq < 10).

Lorsque l'un des effectif théoriques est inférieur à 10 on doit utiliser la formule de la loi binomiale qui donne directement la probabilité P :

$$P_{(k)} = \frac{n!}{k!\ (n-k)!}\ p^k\ q^{n-k} \qquad (A.1.2.1)$$

2. Matrice de contingence.

Prenons comme exemple d'application les séquences dégressives des Ordres pour les ensembles industriels Cjn2, Cjnl et Cbf de la grotte Gatzarria :

Cjn2	S	A	B	SE	E	= P	
	132	93	26	19	0	0	N = 270
Cjnl	S	A	SE	B	E	P	
	70	25	21	6	3	0	N = 125
Cbf	S	SE	E	A	B	P	
	216	76	52	42	8	0	N = 394

2.1. Premier cas : np et nq $\geqslant$ 10.

Pour Cjn2, existe-t-il une discontinuité ou rupture entre S et A ? Avec S = 132, A = 93 et n = 270 on a bien :

$$np = \frac{132+93}{2} = 112{,}5 > 10 \quad \text{et} \quad nq = \frac{138+177}{2} = 157{,}5 > 10$$

On forme le tableau 2x2 où les effectifs (e) des catégories S et A prennent place sur la première ligne et les effectifs complémentaires ($\bar{e}$) dans n (c'est-à-dire non e), sur la seconde ligne :

	S	A	
e	132	93	225
$\bar{e}$	138	177	315
	270	270	540

d'où l'on tire, en applicant les formules (A.1.1.1) et (A.1.2.1) :

$$X^2 = \frac{270\ (132-93)^2}{(132+93)\ (138+177)} = \frac{270.39^2}{225.315} \quad \text{c'est-à-dire :}$$

$$X^2 = 5{,}794 \qquad \text{et } \phi = \sqrt{\frac{5{,}794}{270}} = 0{,}146$$

L'indicateur d'écart $X^2 = 5{,}794$ étant supérieur à $\chi_1^2 = 3{,}841$ il existe une différence significative entre les effectifs des catégories S et A au seuil de signification P = 0,05. Les écarts constatés sont trop grands pour être attribués au hasard. L'indicateur d'écart X^2 mesure le degré de signification.

2.2. Second cas : np ou nq $<$ 10. à la rigueur np ou nq $<$ 5

Pour Cjnl, existe-t-il une discontinuité ou rupture entre B et E ? Avec B = 6, E = 3 et n = 125 on a bien :

$$np = \frac{6+3}{2} = 4{,}5 < 10$$

Il faut dans ce cas calculer la probabilité cumulée d'observer autant de succès ou davantage, ou bien la probabilité d'observer aussi peu de succès ou encore moins, suivant que l'effectif observé s'éloigne de l'effectif théorique en allant vers les grandes ou les petites valeurs.

En appliquant la formule (A.1.2.1) pour l'effectif observé de E = 3, avec n = 125, $p = \frac{np}{n} = \frac{4{,}5}{125}$, $q = \frac{n-np}{n} = \frac{120{,}5}{125}$ avec, itérativement, k = 3,2,1,0 et n-k = 122, 123, 124, 125, on obtient successivement :

$$P(3) = \frac{125!}{3!\ 122!} \left(\frac{4{,}5}{125}\right)^3 \left(\frac{120{,}5}{125}\right)^{122} = 0{,}169$$

$$P(2) = \frac{125!}{2!\ 123!} \left(\frac{4{,}5}{125}\right)^2 \left(\frac{120{,}5}{125}\right)^{123} = 0{,}111$$

$$P(1) = \frac{125!}{1!\ 124!} \left(\frac{4{,}5}{125}\right)^1 \left(\frac{120{,}5}{125}\right)^{124} = 0{,}048$$

$$P(Q) = \frac{125!}{0!\ 125!} \left(\frac{4{,}5}{125}\right)^0 \left(\frac{120{,}5}{125}\right)^{125} = 0{,}010$$

d'où : $P(3-0) = P(3) + P(2) + P(1) + P(0) = 0,338$

La probabilité cumulée 0,338 étant supérieure au seuil de signification 0,05 il n'existe pas de différence significative entre l'effectif observé de E = 3 et l'effectif théorique de 4,5.

De même, en applicant la formule (A.1.2.1) pour l'effectif de B = 6, avec n = 125, $p = \frac{4,5}{125}$, $q = \frac{120,5}{125}$ avec, itérativement, k = 6, 7, 8, 9 et n-k = 119, 118, 117, 116, on obtient la probabilité cumulée :

$$P(6-9) = P(6) + P(7) + P(8) + P(9) = 0,280$$

La probabilité cumulée 0,280 étant supérieure au seuil de signification 0,05 il n'existe pas de différence significative entre l'effectif observé de B = 6 et l'effectif théorique de 4,5.

En conclusion, il n'y a pas de discontinuité entre les catégories B et E.

2.3. Etablissement de la matrice de contingence.

En procédent itérativement (S et A, S et B, S et SE, S et E, S et P, A et B, A et SE, A et E, etc.) on construit une matrice de contingence dans chaque case de laquelle figurent en position supérieure la valeur du X^2 ou, le cas échéant, les valeurs des probabilités calculées à l'aide de la loi binomiale et, en position inférieure, la valeur du ϕ.

A 93	5,794 0,146			Cjn2 N = 270
B 26	50,264 0,431	24,197 0,299		
SE 19	58,694 0,466	30,844 0,338	0,594 0,047	
E=P 0	87,353 0,569	56,174 0,456	13,658 0,255	9,846 0,191
	S 132	A 93	B 26	SE 19

Cjnl

N = 125

A 25	17,190 0,371				
SE 21	20,743 0,407	0,213 0,041			
B 6	38,717 0,557	6,647 0,231	4,671 0,193	(X^2 0,519 P.471) X^2 0,240 ΣP.618	
E 3	43,427 0,584	9,733 0,279	7,467 0,244	P(3-0)= 0,338 P(6-9)= 0,280 0,064	
P 0	48,611 0,624	13,889 0,333	11,463 0,303	P(0) = 0,048 P(6) = 0,049 50 0,157	P(0) = 0,221 P(3) = 0,126 0,110
	S 70	A 25	SE 21	(X^2 3,074 P.081) X^2 2,738 B 6 ΣP.098	E 3 (X^2 1,506 P.220) X^2 0,284 ΣP.347

Cbf

N = 394

SE 76	53,320 0,368				
E 52	76,041 0,439	2,686 0,083			
A 42	87,237 0,471	5,671 0,121	0,604 0,039		
B 8	134,926 0,585	38,455 0,312	17,463 0,211	12,343 0,177	
P 0	148,783 0,615	42,056 0,327	27,837 0,266	22,182 0,237	P(0) = 0,018 9 P(0)8 = 0,029 0,101
	S 216	SE 76	E 52	A 42	B 8 (X^2 4,091 P.044) X^2 3,910 ΣP.048

3. La séquence structurale.

3.1. Les discontinuités ou ruptures.

Lorsque la probabilité correspondant à l'indicateur d'écart X^2, ou la plus petite des probabilités calculées à l'aide de la loi binomiale, est comprise entre 0,05 et 0,01 la différence est dite significative et la discontinuité est marquée par un trait oblique / . Lorsque la probabilité est comprise entre 0,01 et 0,001 la différence est dite très significative et la rupture est marquée par deux traits obliques // . Enfin, lorsque la probabilité est inférieure à 0,001 la différence est dite hautement significative et la rupture est marquée par trois traits obliques ///.

3.2. Le tableau des séquences structurales.

Le tableau des séquences structurales synthétise les résultats obtenus dans les matrices de contingence. On obtient les séquences structurales en complétant les séquences ordonnées par les indications pertinentes de ruptures et, le cas échéant, par des parenthèses mettant en évidence les groupements catégoriels homogènes.

Cjn2	S	/	A	///	B	SE	///	E	=	P	
	132		93		26	19		0		0	N = 270
Cjn1	S	///	A		SE	/	(B	(E)		P)	
	70		25		21		6	3		0	N = 125
Cbf	S	///	(SE		(E)		A)	///	B	/	P
	216		76		52		42		8		0 N = 394

B. ESTIMATIONS DES FREQUENCES THEORIQUES A PARTIR DES FREQUENCES OBSERVEES DANS LA SEQUENCE STRUCTURALE

1. Estimation d'une fréquence théorique à partir d'une fréquence observée f.

On calcule l'intervalle de confiance qui a 95 chances sur 100 de recouvrir la vraie valeur inconnue de p par la formule :

$$f - 1{,}96\sqrt{\frac{f(1-f)}{N}} < p < f + 1{,}96\sqrt{\frac{f(1-f)}{N}} \qquad \text{(B.1.1)}$$

Cette formule n'est valable que si les effectifs théoriques np et nq sont supérieurs à 10 ou, à la rigueur, supérieurs à 5.

Si l'on applique la formule (B.1.1) aux catégories S et A de Cjn2, on obtient avec les données N=270, S=132 et A=93, c'est-à-dire avec les fréquences f(S)=.489 et f(A)=.344 :

pour S $\quad .489 - 1{,}96\sqrt{\frac{.489(1-.489)}{270}} < p < .489 + 1{,}96\sqrt{\frac{.489(1-.489)}{270}}$

$$.429 < p < .549$$

pour A $\quad .344 - 1{,}96\sqrt{\frac{.344(1-.344)}{270}} < p < .344 + 1{,}96\sqrt{\frac{.344(1-.344)}{270}}$

$$.287 < p < .401$$

On remarque que dans le cas de f(S) et de f(A) les intervalles de confiance ne se chevauchent pas :

.287 p(A) .401 .429 p(S) .549

Cette solution de continuité prend la signification d'une rupture.

On notera que ce problème d'estimation ne peut être résolu que si l'on connaît l'effectif N de l'échantillon sur lequel la fréquence f a été déterminée. Isolée, une fréquence f est une information sans valeur : il faut toujours indiquer l'effectif de l'échantillon.

2. Application aux fréquences des séquences structurales.

2.1. Tableau des fréquences observées et des intervalles de confiance.

Cjn2	S	/	A	/	B	SE	/	E	=	P	
	.489 .549 .429		.344 .401 .287		.096 .131 .061	.070 .100 .040		0			N = 270

Cjn1	S	/	A	SE	/	B	E	P	
	.560 .647 .473		.200 .270 .130	.168 .234 .102		.048 .085 .011	.024 .051 .003	0	N = 125

Cbf	S	/	SE	E	A	/	B	P	
	.548 .597 .499		.193 .232 .154	.132 .165 .099	107 .138 .076		.020 .034 .006	0	N = 394

Ainsi, on obtient pour l'intervalle de confiance de B en Cjn1 les valeurs .085 et .011 correspondant aux effectifs théoriques :

a - pour la limite supérieure :
np = 125x0,011 = 10,625 et nq = n(1-p) = 125(1-0,085) = 114,375

b - pour la limite inférieure :
np = 125x0,011 = 1,375 et nq = 125 (1-0,011) = 123,625

Dans le cas de la limite inférieure, du fait que np = 1,375, la valeur ainsi obtenue, plus petite que 5, est douteuse. Il en est de même pour la limite inférieure de l'intervalle de confiance de E en Cjn1 (np = 0,375), ainsi que pour la limite inférieure de l'intervalle de confiance de B en Cbf (np = 2,364).

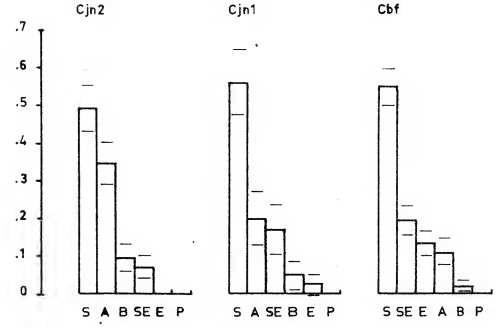
Au tableau des fréquences correspond le tableau des effectifs :

Cjn2	S	/	A	/	B	SE	/	E	=	P
	132 148 116		93 108 78		26 35 17	19 27 11		0		

Cjn1	S	/	A	SE	/	B	E	P
	70 80 60		25 33 17	21 29 13		6 10 2	3 6 0	0

Cbf	S	/	SE	E	A	B	P
	216 235 197		76 91 61	52 65 39	42 54 30	8 13 3	0

2.2. Blocs-fréquences et intervalles de confiance.



C. ANALYSE STRUCTURALE COMPAREE ET ANALYSE DE LA DYNAMIQUE STRUCTURALE

1. Analyse structurale comparée.

L'analyse structurale comparée étudie les rapports entre plusieurs ensembles industriels provenant de sites divers.

Nous prendrons comme exemple d'application les séquences structurales, au niveau des Ordres, des ensembles industriels de La Font Pourquière et des Arènes Candides F4,3,II :

L.F.P.	A	//	S	///	B		P	//	SE	E	
	199		138		22		20		2	2	N = 383
A.C.	A	///	S	///	B		P		SE	E	
	102		38		7		1		1	1	N = 150

1.1. Test d'homogénéité globale.

O étant l'effectif observé et C étant l'effectif théorique calculé, le test d'homogénéité globale se calcule à l'aide des formules :

$$X^2 = \sum \frac{O^2}{C} - N \quad \text{ou} \quad X^2 = \sum \frac{(O-C)^2}{C} \qquad \text{(C.1.1.1)}$$

Le coefficient de contingence quadratique moyenne en dérive selon la formule :

$$C = \sqrt{\frac{X^2}{X^2 + N}} \quad \text{ou} \quad C = \sqrt{\frac{\phi^2}{\phi^2 + 1}} \qquad (C.1.1.2)$$

Testons l'homogénéité des deux ensembles industriels de La Font Pourquière et des Arènes Candides F4,3,II. Nous obtenons :

$X^2 = 12,75$ $\qquad 0,01 > P > 0,001 \qquad C = 0,153$

Les écarts constatés sont trop grands pour être attribué au hasard. L'hypothèse nulle de l'homogénéité ne pouvant être retenue, on admet l'hypothèse alternative de l'inhomogénéité.

1.2. Recherche des différences catégorielles significatives.

1.2.1. Premier cas: les effectifs calculés sont supérieurs à 5.

On utilise la formule exclusive des tableaux 2x2 :

$$X^2 = \frac{N\ (ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (C.1.2.1.1)$$

d'où l'on déduit le coefficient de corrélation de points :

$$\phi = \sqrt{\frac{X^2}{N}} \qquad (C.1.2.1.2)$$

Testons l'homogénéité des catégories A et non A ($\bar{A}$) :

	J_1	J_2	
I_1	a	b	a+b
I_2	c	d	c+d
	a+c	b+d	N

	LFP	AC	
A	199	102	301
$\bar{A}$	184	48	232
	383	150	533

On obtient: $X^2 = 36,013$; $\phi = 0,260$; $P < 0,001$ inhomogénéité

Pour S et $\bar{S}$: $X^2 = 5,578$ $\phi = 0,102$ $0,05 > P > 0,01$ inhomogénéité
Pour B et $\bar{B}$: $X^2 = 0,243$ $\phi = 0,021$ $0,50 > P > 0,20$ homogénéité
Pour P et $\bar{P}$: $X^2 = 5,910$ $\phi = 0,105$ $0,05 > P > 0,01$ inhomogénéité

1.2.2. Second cas: les effectifs calculés sont inférieurs à 5.

On utilise la "méthode exacte de Fisher" en calculant la probabilité cumulée d'observer un effectif a aussi grand ou encore plus grand, ou aussi petit ou encore plus petit relativement à l'effectif théorique calculé. On applique la formule dérivée de la loi hypergéométrique :

$$P(a) = \frac{(a+b)!\ (a+c)!\ (b+d)!\ (c+d)!}{N!\ a!\ b!\ c!\ d!} \qquad (C.1.2.2)$$

Ainsi, pour SE et $\overline{SE}$ on obtient :

	LFP	AC	
SE	2	1	3
$\overline{SE}$	381	149	530
	383	150	533

$$P(2) = \frac{530!\ 383!\ 150!\ 3!}{533!\ 381!\ 149!\ 2!\ 1!}$$

que l'on simplifie comme suit :

$$P(2) = \frac{(382 \cdot 383)\ (150)\ (3)}{(531 \cdot 532 \cdot 533)} = 0,437$$

	LFP	AC	
SE	1	2	3
$\overline{SE}$	382	148	530
	383	150	533

$$P(1) = \frac{530!\ 383!\ 150!\ 3!}{533!\ 382!\ 148!\ 2!\ 1!}$$

$$P(1) = \frac{(383)(149 \cdot 150)\ (3)}{(531 \cdot 532 \cdot 533)} = 0,171$$

	LFP	AC	
SE	0	3	3
$\overline{SE}$	383	147	530
	383	150	533

$$P(0) = \frac{530!\ 383!\ 150!\ 3!\ 0!}{533!\ 383!\ 147!\ 3!\ 0!}$$

$$P(0) = \frac{148 \cdot 149 \cdot 150}{531 \cdot 532 \cdot 533} = 0,022$$

$P(2) + P(1) + P(0) = 0,63$ homogénéité

Pour E et $\overline{E}$ on obtient de même :

$P(2) + P(1) + P(0) = 0,63$ homogénéité

1.2.3. Tableau d'analyse structurale comparée.

	L.F.P.	A.C.	
A	.519	.680	différence hautement significative positive pour AC
S	.360	.253	différence significative positive pour LFP
B	.057	.047	différence non significative
P	.052	.007	différence significative positive pour FLP
SE	.005	.007	différence non significative
E	.005	.007	différence non significative

2. Analyse de la dynamique structurale.

L'analyse de la dynamique structurale étudie les rapports entre les ensembles industriels d'une séquence stratigraphique.

Nous prendrons pour exemple d'application la séquence stratigraphique des ensembles industriels Cjn2, Cjnl et Cbf de la grotte Gatzarria, au niveau des Ordres.

2.1. Tests d'homogénéité globale.

Pour les ensembles industriels Cjn2, Cjnl et Cbf, on obtient :

$X^2 = 125,365$ $C = 0,370$ $P < 0,001$ inhomogénéité

Pour les ensembles industriels Cjn2 et Cjnl, on obtient :

$X^2 = 15,311$ $C = 0,491$ $0,001 < P < 0,01$ inhomogénéité

Pour les ensembles industriels Cjnl et Cbf, on obtient :

$X^2 = 19,815$ $C = 0,192$ $P < 0,001$ inhomogénéité

2.2. Recherche des différences catégorielles globales significatives.

On utilise les formules (C.1.1.1) et (C.1.1.2).
Pour les ensembles industriels Cjn2, Cjnl et Cbf, on obtient :

S et $\bar{S}$	: X^2 = 2,808	C = 0,060	0,20 < P < 0,30	homogénéité
A et $\bar{A}$	: X^2 = 56,076	C = 0,237	P < 0,001	inhomogénéité
SE et $\bar{SE}$	: X^2 = 19,700	C = 0,156	P < 0,001	inhomogénéité
E et $\bar{E}$	: X^2 = 47,800	C = 0,257	P < 0,001	inhomogénéité
B et $\bar{B}$	: X^2 = 19,237	C = 0,154	P < 0,001	inhomogénéité

2.3. Recherche des différences catégorielles significatives en articulation stratigraphique.

On utilise les formules (C.1.2.1.1), (C.1.2.1.2) et (C.1.2.2).
Pour les ensembles industriels Cjn2 et Cjnl, on obtient :

S et $\bar{S}$	: X^2 = 1,729	ϕ = 0,066	0,10 < P < 0,20	homogénéité
A et $\bar{A}$	: X^2 = 8,510	ϕ = 0,147	0,001 < P < 0,01	inhomogénéité
SE et $\bar{SE}$	: X^2 = 8,948	ϕ = 0,151	0,001 < P < 0,01	inhomogénéité
E et $\bar{E}$	: P(3-0)=0,031	ϕ = 0,129	P = 0,031	inhomogénéité
B et $\bar{B}$	: X^2 = 2,677	ϕ = 0,082	0,10 < P < 0,20	homogénéité

Pour les ensembles industriels Cjnl et Cbf, on obtient :

S et $\bar{S}$	: X^2 = 0,053	ϕ = 0,010	0,80 < P < 0,90	homogénéité
A et $\bar{A}$	: X^2 = 7,363	ϕ = 0,119	0,001 < P < 0,01	inhomogénéité
SE et $\bar{SE}$	: X^2 = 0,387	ϕ = 0,027	0,50 < P < 0,70	homogénéité
E et $\bar{E}$	: X^2 = 11,678	ϕ = 0,150	P < 0,001	inhomogénéité
B et $\bar{B}$	: X^2 = 2,773	ϕ = 0,073	0,05 < P < 0,10	homogénéité

2.4. Tableau de la dynamique structurale.

Le tableau de la dynamique structurale synthétise, au niveau structural considéré, les résultats obtenus par la recherche des différences catégorielles significatives globales (& 2.2) et en articulation stratigraphique (& 2.3). Les ensembles industriels sont rangés selon l'ordre stratigraphique, les catégories selon l'ordre des fréquences théoriques (p). On mentionne, en outre, les valeurs des amplitudes des fréquences catégorielles (A) et les valeurs des amplitudes des fréquences extrêmes (Ae). Les valeurs A et Ae permettent de caractériser comme suit les mouvements déduits de leurs différences :

Ae = A = 0 ..... stabilité
Ae = A ≠ 0 ..... progression ou dégression
Ae ≠ A et Ae = 0 ..... oscillation
Ae ≠ A et 2Ae < A ..... oscillation progressive ou dégressive
Ae ≠ A et 2Ae > A ..... progression ou dégression oscillante

Dressons le tableau de dynamique structurale au niveau des Ordres de la séquence stratigraphique Cjn2, Cjnl et Cbf de la grotte Catzarria.

	Cjn2	Cjnl	Cbf	p	A	Ae	Dynamique
S	.489	.560	.548	.530	.071	.059	progression oscillante
A	.344 //	.200 //	.107	.203	.237	.237	dégression haut.signif.
SE	.070 //	.168	.193	.147	.123	.123	progression haut.signif
E	0 /	.024 ///	.132	.070	.132	.132	progression haut.signif
B	.096	.048	.020	.051	.076	.076	dégression haut.signif.
P	0	0	0				stabilité absence

N.B. Les recherches concernant le "Critère de contingence dans la séquence structurale" ainsi que les "Estimations des fréquences théoriques à partir des fréquence observées dans la séquence structurale" n'ont pas été exposées durant le séminaire du mois d'Août 1975. Elles feront partie du programme du séminaire du mois d'Août 1976.

DISTANCE DU KHI 2 ET ALGORITHMES DE CLASSIFICATION HIERARCHIQUE

Georges Laplace

On se propose de classer selon leurs ressemblances :

a - soit des séries ou ensembles industriels considérés à un niveau structural donné ;

b - soit les catégories composantes du niveau structural considéré dans les séries ou ensembles industriels donnés.

La distance choisie entre les séries ou entre les catégories est appelée "distance du Khi 2". Deux algorithmes, ou procédés de calcul, de classification hiérarchique indicée sont proposés : l'un consistant à rendre ultramétrique la distance d, l'autre procédant par réduction.

A. DISTANCE DU KHI 2.

1. Formules de la distance du Khi 2.

Soit le tableau de contingence r x k, c'est-à-dire comportant r lignes et k colonnes :

		Catégories j	Catégories j'		
Séries	i	n_{ij}	$n_{ij'}$		$n_{i.}$
	i'	$n_{i'j}$	$n_{i'j'}$		$n_{i'.}$
		$n_{.j}$	$n_{.j'}$		$n_{..}$

(A.1.1)

Dans ce tableau de données brutes, comportant k catégories et r séries :

a) n_{ij} exprime le nombre d'observations présentant à la fois la modalité i de la variable "séries" et la modalité j de la variable "catégories" ;

b) $n_{i.}$ représente l'effectif de la série i, c'est-à-dire la somme le long de i, j variant de 1 à k, soit :

$$n_{i.} = \sum_{j=1}^{k} n_{ij}$$

c) $n_{.j}$ représente l'effectif de la catégorie j, c'est-à-dire la somme le long de j, i variant de 1 à r, soit :

$$n_{.j} = \sum_{i=1}^{r} n_{ij}$$

d) $n_{..}$ représente l'effectif total, soit :

$$n_{..} = \sum_{j=1}^{k} \sum_{i=1}^{r} n_{ij}$$

A partir de ce tableau on construit le tableau des fréquences :

Catégories

Séries		j	j'		
i		f_{ij}	$f_{ij'}$		$f_{i.}$
i'		$f_{i'j}$	$f_{i'j'}$		$f_{i'.}$
		$f_{.j}$	$f_{.j'}$		1

(A.1.2)

Le fait d'avoir choisi des "profils", ou fréquences conditionnelles, conduit à adopter une distance différente de la distance euclidienne usuelle, laquelle se calcule comme suit :

$$d^2(i, i') = \sum_{j=1}^{k} \left(\frac{n_{ij}}{n_{i.}} - \frac{n_{i'j}}{n_{i'.}}\right)^2$$

ou $$d^2(i, i') = \sum_{j=1}^{k} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}}\right)^2$$

car $f_{ij} = \frac{n_{ij}}{n_{..}}$, $f_{i'j} = \frac{n_{i'j}}{n_{..}}$, $f_{i.} = \frac{n_{i.}}{n_{..}}$ et $f_{i'.} = \frac{n_{i'.}}{n_{..}}$

et, de façon symétrique :

$$d^2(j, j') = \sum_{i=1}^{r} \left(\frac{n_{ij}}{n_{\bullet j}} - \frac{n_{ij'}}{n_{\bullet j'}}\right)^2$$

ou $$d^2(j, j') = \sum_{i=1}^{r} \left(\frac{f_{ij}}{f_{\bullet j}} - \frac{f_{ij'}}{f_{\bullet j'}}\right)^2$$

La distance du Khi 2 ne diffère en fait de la métrique euclidienne usuelle que par la pondération de chaque carré par les inverses des fréquences correspondant à chaque terme.

La distance du Khi 2 entre deux modalités sérielles i et i' sera donnée par les formules :

$$d^2(i, i') = \sum_{j=1}^{k} \frac{n_{\bullet\bullet}}{n_{\bullet j}} \left(\frac{n_{ij}}{n_{i\bullet}} - \frac{n_{i'j}}{n_{i'\bullet}}\right)^2 \qquad (A.1.3)$$

ou $$d^2(i, i') = \sum_{j=1}^{k} \frac{1}{f_{\bullet j}} \left(\frac{f_{ij}}{f_{i\bullet}} - \frac{f_{i'j}}{f_{i'\bullet}}\right)^2 \qquad (A.1.4)$$

De façon symétrique, la distance du Khi 2 entre les deux modalités catégorielles j et j' vaut :

$$d^2(j, j') = \sum_{i=1}^{r} \frac{n_{\bullet\bullet}}{n_{i\bullet}} \left(\frac{n_{ij}}{n_{\bullet j}} - \frac{n_{ij'}}{n_{\bullet j'}}\right)^2 \qquad (A.1.5)$$

ou $$d^2(j, j') = \sum_{i=1}^{r} \frac{1}{f_{i\bullet}} \left(\frac{f_{ij}}{f_{\bullet j}} - \frac{f_{ij'}}{f_{\bullet j'}}\right)^2 \qquad (A.1.6)$$

2. Exemple d'application des formules de la distance du Khi 2.

Prenons pour exemple d'application de la formule de la distance du Khi 2 les effectifs des trois classes de Grattoirs (G.f = grattoirs frontaux ; G.m = grattoirs à museau ; G.c = grattoirs carénés) dans les séries ou ensembles industriels Cjn2, Cjn1, Cbf et Cb de la grotte Gatzarria. On construit le tableau de contingence (A.1.1) :

	G.f	G.m	G.c	
Cjn2	5	4	14	23
Cjn1	10	8	20	38
Cbf	33	21	61	115
Cb	41	27	72	140
	89	60	167	316

Si l'on se propose de calculer la distance entre les quatre séries ou ensembles industriels, on doit utiliser la formule (A.1.3). Par contre, si l'on avait voulu calculer les distances entre les trois catégories de grattoirs on aurait utilisé la formule symétrique (A.1.5).

Calculons la distance du Khi 2 entre Cjn2 et Cjnl :

$$d^2(\text{Cjn2,Cjnl}) = \frac{316}{89}\left(\frac{5}{23} - \frac{10}{38}\right)^2 + \frac{316}{60}\left(\frac{4}{23} - \frac{8}{38}\right)^2 + \frac{316}{167}\left(\frac{14}{23} - \frac{20}{38}\right)^2$$

$$= 3{,}551\ (.217 - .263)^2 + 5{,}267\ (.174 - .211)^2 + 1{,}892\ (.609 - .526)^2$$

$$= 0{,}028$$

d'où $\quad d(\text{Cjn2, Cjnl}) = \sqrt{0{,}028} = 0{,}167$

Pour faciliter les calculs on dresse le tableau pertinent des fréquences utilisées dans la formule (A.1.3) :

		j	j'		
i		$\frac{n_{ij}}{n_{i\circ}}$	$\frac{n_{ij'}}{n_{i\circ}}$		1
i'		$\frac{n_{i'j}}{n_{i'\circ}}$	$\frac{n_{i'j'}}{n_{i'\circ}}$		1
1/f		$\frac{n_{\circ\circ}}{n_{\cdot j}}$	$\frac{n_{\circ\circ}}{n_{\cdot j'}}$		

et on obtient :

	G.f	G.m	G.c	
Cjn2	.217	.174	.609	1
Cjnl	.263	.211	.526	1
Cbf	.287	.183	.530	1
Cb	.293	.193	.514	1
1/f	3,551	5,267	1,892	

A partir de ce tableau on calcule itérativement les distances entre les quatre séries ou ensembles industriels :

$d^2(\text{Cjn2, Cjnl}) = 0{,}028$ d'où $d(\text{Cjn2, Cjnl}) = 0{,}167$
$d^2(\text{Cjn2, Cbf}) = 0{,}030$ d'où $d(\text{Cjn2, Cbf}) = 0{,}172$
$d^2(\text{Cjn2, Cb}) = 0{,}039$ d'où $d(\text{Cjn2, Cb}) = 0{,}199$
$d^2(\text{Cjnl, Cbf}) = 0{,}006$ d'où $d(\text{Cjnl, Cbf}) = 0{,}079$
$d^2(\text{Cjnl, Cb}) = 0{,}005$ d'où $d(\text{Cjnl, Cb}) = 0{,}072$
$d^2(\text{Cbf, Cb}) = 0{,}001$ d'où $d(\text{Cbf, Cb}) = 0{,}034$

B. ALGORITHME UTILISANT UNE DISTANCE ULTRAMETRIQUE.

Cet algorithme concerne le passage d'une matrice de distance à une matrice ultramétrique. Plusieurs procédés peuvent être utilisés. Nous en présentons trois :

a - celui de l'ultramétrique supérieure minimale ;
b - celui de l'ultramétrique inférieure maximale ;
c - celui de l'ultramétrique moyenne.

Soit E un ensemble de n objets, muni d'un indice de "distance" d, ici la distance du Khi 2. On aurait pu prendre un indice de similarité mais, pour rester homogène avec le sens de variation d'une distance, on a préféré considérer un indice qui varie en sens contraire d'une similarité. Un tel indice est appelé quelquefois indice de dissimilarité. Si d n'est pas ultramétrique, il n'induit pas une seule hiérarchie indicée. Le choix d'une formule pour calculer à chaque étape les dissimilarités entre parties de E s'impose. Quand on choisit une telle formule, on construit très simplement une hiérarchie indicée correspondante, c'est-à-dire un dendrogramme niveau par niveau. En haut, on place les n points de E. Au niveau 1, on agrège les deux points les plus proches et on calcule la dissimilarité des n - 2 autres points à cette partie à l'aide de la formule, et ainsi de suite. On s'arrête au niveau n - 1.

Remarque : cet algorithme est ascendant puisqu'il procède par regroupement des séries ou des catégories.

1. Construction d'une ultramétrique supérieure minimale.

L'algorithme proposé revient à rendre isocèle tout triangle de l'indice initial en lui donnant pour base son plus petit côté et pour longueur des deux côtés égaux celle de son plus grand côté. On construit ainsi une hiérarchie indicée, et l'ultramétrique correspondante u_1 est "supérieure" à la dissimilarité initiale. On démontre que cette ultramétrique est minimale dans l'ensemble des ultramétriques supérieures à d.

On dispose les résultats obtenus par le calcul de la distance du Khi 2 de manière à former une matrice des distances entre les quatre séries ou ensembles industriels :

d	Cjn2	Cjnl	Cbf	Cb
Cjn2	0	0,167	0,177	0,199
Cjnl		0	0,079	0,072
Cbf			0	0,034
Cb				0

On constate que la distance la plus courte est de 0,034 entre

les séries Cb et Cbf. En conséquence, on agrège Cb et Cbf et on calcule, pour obtenir une nouvelle matrice des distances ou tableau de dissimilarité, les distances entre la nouvelle série (Cb,Cbf) et les séries Cjn2 et Cjnl.

Soit le triangle Cb-Cbf-Cjn2. On obtient les longueurs : 0,034 pour le côté Cb-Cbf, 0,199 pour le côté Cb-Cjn2 et 0,177 pour le côté Cbf-Cjn2. On choisira donc 0,034, longueur du plus petit côté, pour longueur de la base et 0,199, longueur du plus grand côté, pour longueur des deux côtés égaux. Ainsi, la distance ultramétrique entre (Cb,Cbf) et Cjn2 sera de 0,199.

De même, si l'on considère le triangle Cb-Cbf-Cjnl, on obtient les longueurs : 0,034 pour le côté Cb-Cbf, 0,072 pour le côté Cb-Cjnl et 0,079 pour le côté Cbf-Cjnl. Par conséquent, le triangle "isocèle pointu" aura une base de longueur 0,034 et deux côtés égaux de longueur 0,079. Ainsi, la distance ultramétrique entre (Cb,Cbf) et Cjnl sera de 0,079. La distance entre Cjn2 et Cjnl, soit 0,167, ne variant pas on construit sur l'ensemble de ces données la nouvelle matrice des distances:

1re étape	(Cb,Cbf)	Cjn2	Cjnl
(Cb,Cbf)	0	0,199	0,079
Cjn2		0	0,167
Cjnl			0

Procédant comme précédemment, on constate que la distance la plus courte est de 0,079 entre la série (Cb,Cbf) et la série Cjnl.

On agrège (Cb,Cbf) et Cjnl et on calcule, pour obtenir une nouvelle matrice des distances, la distance entre la nouvelle série (Cb,Cbf,Cjnl) et la série Cjn2.

On rend "isocèle pointu" le triangle (Cb,Cbf)-Cjnl-Cjn2 de côtés (Cb,Cbf)-Cjnl = 0,079, (Cb,Cbf)-Cjn2 = 0,199 et Cjnl-Cjn2 = 0,167. On en déduit la distance ultramétrique entre la série (Cb, Cbf,Cjnl) et la série Cjn2, soit 0,199. On construit la dernière matrice des distances :

2e étape	(Cb,Cbf, Cjnl)	Cjn2
(Cb,Cbf,Cjnl)	0	0,199
Cjn2		0

Nous passons donc de la matrice des distances à une matrice ultramétrique u_1 :

u_1	Cjn2	Cjnl	Cbf	Cb
Cjn2	0	0,199	0,199	0,199
Cjnl		0	0,079	0,079
Cbf			0	0,034
Cb				0

Remarque : on constate que toutes les distances figurant dans ce tableau sont "ultramétriques", c'est-à-dire qu'elles vérifient la condition dite de Bourbaki :

"Pour tout triplet $(X_h, X_i, X_e,)$ on a :

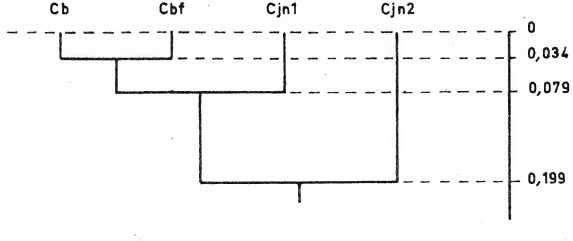
$$d(X_h, X_i) \leqslant \text{SUP}\ d(X_h, X_e'), d(X_e, X_i)"$$

Chaque fois que l'on considère trois séries (ou catégories) la distance entre deux séries (ou catégories) est toujours inférieure ou égale à la plus grande des distances de ces deux séries (ou catégories) à la troisième, ce qui revient à dire que tous les triangles ayant pour sommets les séries (ou catégories) sont isocèles ou équilatéraux.

On peut maintenant dessiner le dendrogramme. Il se présente comme un arbre de classification associé à une échelle de distance ultramétrique et figure, de ce fait, une hiérarchie stratifiée indicée.

feuilles

Il est évident que deux séries (ou catégories) figurant les branches se ressemblent d'autant plus que le premier noeud qui les relie est plus élevé dans l'arbre.



2. Construction d'une ultramétrique inférieure maximale.

L'algorithme proposé revient à rendre isocèle tout triangle de l'indice initial en lui donnant pour base son plus petit côté et en réduisant le plus grand côté à la longueur du côté immédiatement inférieur. On obtient une hiérarchie différente de la précédente et l'ultramétrique correspondante u_2 n'est donc pas équivalente à u_1 : elle est inférieure à d et on démontre qu'elle est maximale dans l'ensemble des ultramétriques inférieures à d.

Reprenons la matrice des distances :

d	Cjn2	Cjnl	Cbf	Cb
Cjn2	0	0,167	0,177	0,199
Cjnl		0	0,079	0,072
Cbf			0	0,034
Cb				0

On obtient successivement les matrices des distances :

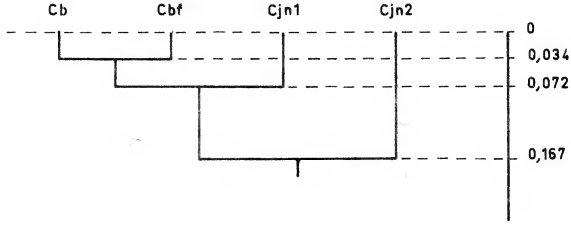
1re étape	(Cb,Cbf)	Cjn2	Cjnl
(Cb,Cbf)	0	0,177	0,072
Cjn2		0	0,167
Cjnl			0

2e étape	(Cb,Cbf,Cjnl)	Cjn2
(Cb,Cbf,Cjnl)	0	0,167
Cjn2		0

la matrice ultramétrique :

u_2	Cjn2	Cjnl	Cbf	Cb
Cjn2	0	0,167	0,167	0,167
Cjnl		0	0,072	0,072
Cbf			0	0,034
Cb				0

et, enfin, le dendrogramme :



3. Construction d'une ultramétrique moyenne.

L'algorithme proposé revient à rendre isocèle tout triangle de l'indice initial en lui donnant pour base son plus petit côté et pour longueur des deux côtés égaux la moyenne des longueurs des deux autres côtés.

Reprenons la matrice des distances :

d	Cjn2	Cjnl	Cbf	Cb
Cjn2	0	0,167	0,177	0,199
Cjnl		0	0,079	0,072
Cbf			0	0,034
Cb				0

On obtient successivement les matrices des distances :

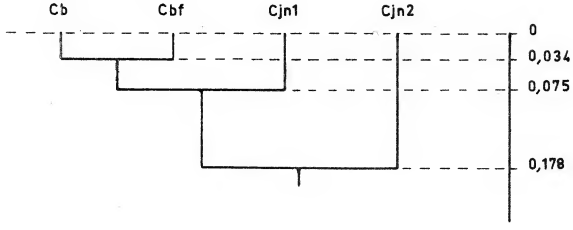
1re étape	(Cb,Cbf	Cjn2	Cjnl
(Cb,Cbf)	0	0,188	0,075
Cjn2		0	0,167
Cjnl			0

2e étape	(Cb,Cbf,Cjnl)	Cjn2
(Cb,Cbf,Cjnl)	0	0,178
Cjn2		0

la matrice ultramétrique :

u_3	Cjn2	Cjnl	Cbf	Cb
Cjn2	0	0,178	0,178	0,178
Cjnl		0	0,075	0,075
Cbf			0	0,034
Cb				0

et, enfin, le dendrogramme :



C. ALGORITHME PROCEDANT PAR REDUCTIONS.

Cet algorithme procède par regroupements successifs de façon à réduire progressivement les dimensions de la matrice de départ.

Remarque. Cet algorithme est ascendant puisqu'il procède par regroupement des séries ou des catégories.

1. L'algorithme.

La démarche est exposée à partir du tableau pertinent des fréquences (A.1.7) utilisées dans la formule (A.1.3) pour les séries ou ensembles industriels Cjn2, Cjnl, Cbf et Cb de la grotte Gatzarria considérés au niveau structural des classes de Grattoirs :

	G.f	G.m	G.c	
Cjn2	.217	.174	.609	1
Cjnl	.263	.211	.526	1
Cbf	.287	.183	.530	1
Cb	.293	.193	.514	1
1/f	3,551	5,267	1,892	

Les distances du Khi 2 calculées entre les quatre séries ou ensembles industriels sont reportées dans une matrice des distances ou tableau de dissimilarité :

d	Cjn2	Cjnl	Cbf	Cb
Cjn2	0	0,167	0,177	0,199
Cjnl		0	0,079	0,072
Cbf			0	0,034
Cb				0

On constate que la distance la plus courte est de 0,034 entre les séries Cb et Cbf. En conséquence, on agrège les séries Cb et Cbf, soit (Cb,Cbf), pour former avec les séries Cjn2 et Cjnl un nouveau tableau des fréquences :

	G.f	G.m	G.c	
(Cb,Cbf)	.290	.188	.522	1
Cjn2	.217	.174	.609	1
Cjnl	.263	.211	.526	1
1/f	3,551	5,267	1,892	

Calculons les distances du Khi 2 entre les trois séries Cjn2, Cjnl et (Cb,Cbf) :

d^2 (Cjn2,Cjnl) = 0,028 d'où d (Cjn2,Cjnl) = 0,167

d^2 ((Cb,Cbf),Cjn2) = 0,034 d'où d ((Cb,Cbf),Cjn2) = 0,185

d^2 ((Cb,Cbf),Cjnl) = 0,005 d'où d ((Cb,Cbf),Cjnl) = 0,074

La matrice se réduit comme suit :

1re étape	(Cb,Cbf)	Cjn2	Cjnl
(Cb,Cbf)	0	0,185	0,074
Cjn2		0	0,167
Cjnl			0

On constate que la distance la plus courte est de 0,074 entre (Cb,Cbf) et Cjnl. En conséquence, on agrège les séries (Cb,Cbf) et

Cjnl, soit (Cb, Cbf, Cjnl), pour former avec la série Cjn2 un nouveau tableau de fréquences :

	G.f	G.m	G.c	
(Cb,Cbf,Cjnl)	.287	.191	.522	1
Cjn2	.217	.174	.609	1
1/f	3,551	5,267	1,892	

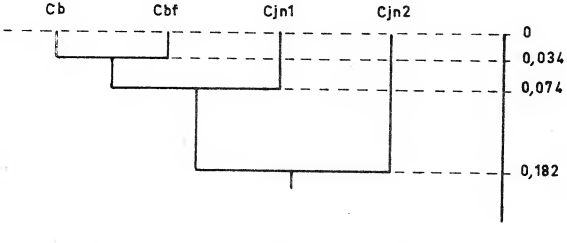
Calculons les distances du Khi 2 entre les deux séries (Cb,Cbf, Cjnl) et Cjn2 :

$d^2((Cb,Cbf,Cjnl),Cjn2) = 0,033$ d'où $d((Cb,Cbf,Cjnl),Cjn2) = 0,182$

La matrice se réduit enfin comme suit :

2e étape	(Cb,Cbf,Cjnl)	Cjn2
(Cb,Cbf,Cjnl)	0	0,182
Cjn2		0

Nous pouvons maintenant dessiner le dendrogramme, c'est-à-dire un arbre de classification qui,muni d'une échelle de distance, figure une hiérarchie indicée.



2. Autre exemple d'application de l'algorithme.

On se propose de classer hiérarchiquement les catégories d'Ordres (S = Simples; A = Abrupts; SE = Surélevés; E = Ecaillés; B = Burins et P = Plans) dans les trois séries ou ensembles industriels Cjn2, Cjnl et Cbf de la grotte Gatzarria.

On compose avec ces données le tableau de contingence à r séries et k catégories (A.1.1) :

	S	A	SE	E	B	P	
Cjn2	132	93	19	0	26	0	270
Cjn1	70	25	21	3	6	0	125
Cbf	216	42	76	52	8	0	394
	418	160	116	55	40	0	789

Pour faciliter le calcul de la distance du Khi 2 on dresse le tableau pertinent des fréquences utilisées dans la formule (A.1.5):

$$d^2(j,j') = \sum_{i=1}^{k} \frac{n_{..}}{n_{i.}} \left(\frac{n_{ij}}{n_{.j}} - \frac{n_{ij'}}{n_{.j'}} \right)^2$$

		j	j'		
i		$\frac{n_{ij}}{n_{.j}}$	$\frac{n_{ij'}}{n_{.j'}}$		$\frac{n_{..}}{n_{i.}}$
i'		$\frac{n_{i'j}}{n_{.j}}$	$\frac{n_{i'j'}}{n_{.j'}}$		$\frac{n_{..}}{n_{i'.}}$
		1	1		

c'est-à-dire :

	S	A	SE	E	B	1/f
Cjn2	.316	.581	.164	.	.650	2,922
Cjn1	.167	.156	.181	.055	.150	6,312
Cbf	.517	.263	.655	.945	.200	2,003
	1.	1.	1.	1.	1.	

La catégorie P disparaît car le quotient $\frac{n_{ij}}{n_{.j}} = \frac{f_{ij}}{f_{.j}} = \frac{0}{0}$

est une forme indéterminée. Par contre, il en serait tout autrement si le numérateur étant égal à zéro le dénominateur était différent de zéro.

Les distances du Khi 2 calculées entre les cinq catégories sont reportées dans une matrice des distances ou tableau de dissimilarité:

	S	A	SE	E	B
S	0	0,579	0,327	0,859	0,727
A		0	0,905	1,408	0,149
SE			0	0,584	1,054
E				0	1,550
B					0

On constate que la distance la plus courte est de 0,149 entre les catégories A et B. En conséquence, on agrège ces deux catégories, soit (A,B), pour former avec les catégories S,SE,E et B un nouveau tableau des fréquences :

	(A,B)	S	SE	E	1/f
Cjn2	.595	.316	.164	.	2,922
Cjn1	.155	.167	.181	.055	6,312
Cbf	.250	.517	.655	.945	2,003
	1.	1.	1.	1.	

On calcule les nouvelles distances du Khi 2 et on les reporte dans une nouvelle matrice :

	(A,B)	S	SE	E
(A,B)	0	0,609	0,936	1,437
S		0	0,327	0,859
SE			0	0,589
E				0

On constate que la distance la plus courte est de 0,327 entre les catégories S et SE. En conséquence, on agrège ces catégories,

soit (S,SE), pour former avec les catégories (A,B) et E un nouveau tableau des fréquences :

	(A,B)	(S,SE)	E	1/f
Cjn2	.595	.282	.	2,922
Cjnl	.155	.170	.055	6,312
Cbf	.250	.547	.945	2,003
	1.	1.	1.	

On calcule les nouvelles distances du Khi 2 et on les reporte dans une nouvelle matrice des distances :

	(A,B)	(S,SE)	E
(A,B)	0	0,681	1,437
(S,SE)		0	0,797
E			0

On constate que la distance la plus courte est de 0,681 entre (S,SE) et (A,B). On agrège donc ces catégories, soit (A,B,S,SE) pour former avec E un nouveau tableau des fréquences :

	(A,B,S,SE)	E	1/f
Cjn2	.368	.	2,922
Cjnl	.166	.055	6,312
Cbf	.466	.945	2,003
	1.	1.	

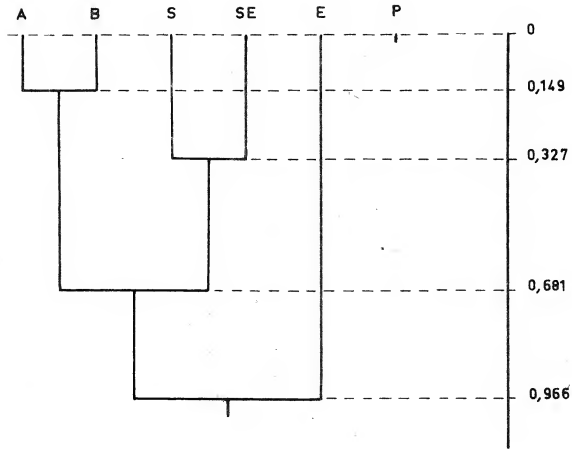
On calcule les nouvelles distances du Khi 2 et nous les reportons dans une dernière matrice des distances :

	(A,B,S,SE)	E
(A,B,S,SE)	0	0,966
E		0

L'ultime distance est de 0,966, entre les catégories (A,B,S,SE) et E.

Nous pouvons maintenant dessiner le dendrogramme qui exprime les analogies évolutives des catégories en une classification hié-

rarchisée et indicée.



BIBLIOGRAPHIE

P. BERTIER, J.M. BOUROCHE - Analyse des données multidimensionnelles. Presses Universitaires de France, 1975.

N.B. Cette étude résulte des recherches les plus récentes. Elle fera l'objet d'un exposé au cours du séminaire du mois d'Août 1976.

LA FICHE DE TYPOLOGIE ANALYTIQUE

Robert Vilain

L'utilisation de la méthode de Typologie Analytique de G.Laplace nous avait conduit, avec S.Morelon, à l'élaboration d'une fiche d'analyse permettant la transcription rapide des caractères morphotechniques observés sur chaque outil étudié. La fiche presentée tient compte des dernières données de la typologie analytique et des observation que nous avons pu faire au cours de son utilisation.

Il ne me semble pas utile de commenter la fiche ici reproduite (Fig.1) si ce n'est pour préciser que :

- il nous est apparu souhaitable de laisser un blanc dans le cartouche de l'intitulé; on peut ainsi noter des renseignements annexes n'ayant qu'une importance secondaire, voire personnels.
Par exemple : la date d'analyse, le nom du chercheur, ou la référence bibliographique.

- la colonne : observations - mesures, s'est révélée indispensable pour noter les dimensions, les cassures affectant telle ou telle autre partie de la pièce, les fragments divers: distaux, proximaux, médians.

- les sigles conventionnels de la méthode marquant : la discontinuité _ _ , l'opposition · , la continuité _ , la complémentarité / ou la composition latéro transversale + , prennent place dans la colonne: retouches complémentaires.

- l'indication tendance ( ), peut s'inscrire dans les autres colonnes; par exemple une retouche peut être simple, une croix est alors inscrite dans la colonne afférente, mais s'il est estimé que l'ampleur de cette retouche et son angle la placent à la limite de la retouche plate, il suffira de noter le sigle ( ) dans la colonne: plate, pour reproduire cette notion d'incertitude. Il en est bien entendu de même pour les caractères d'amplitude : profonde, margina-

le; de variété : envahissante, couvrante, sommaire, scalariforme, laminaire.

- dans le cas où une retouche est très marginale ou liminale, mm, c'est le cas des retouches fortuites d'utilisation, la croix est doublée dans la colonne: marginale.

Cette fiche ne comporte pas de colonne pour de plus grandes précisions; d'une part la place manquait et, d'autre part, la multiplication des rubriques qui peuvent varier, grâce à la souplesse de la méthode, dans de grandes proportions par simple combinaison des caractères, eut rendu cet instrument de travail peu maniable. La complexité n'engendrant pas forcément l'efficacité. Par exemple, pour noter qu'un retouche n'interesse que la partie distale du bord dextre de l'outil observé, il suffit de relier par un trait continu les croix inscrites respectivement dans les colonnes: dextre et distale.

Dans le cas d'une retouche occupant la totalité d'un bord, il suffit de noter simplement la localisation dextre ou senestre, sans autre précision.

Il en est de même pour la précision: rectiligne; elle est entendue s'il n'est rien noté dans les colonnes convexe ou concave.

Ce type de fiche d'analyse typologique permet de travailler à deux très rapidement, la lecture de l'analyste étant traduite directement en un résumé codé. Seul, le travail en est également de beaucoup facilité.

Enfin, un autre avantage de ce type de fiche réside dans le fait que l'on peut extraire, pour chaque type d'outil, le ou les caractères les plus fréquents ou, au contraire, les plus rares, que il s'agisse de tendance, de localisation, d'ampleur de la retouche.

Pour chaque type d'outil une ou plusieurs fiches sont employées. Dans le cas de leur multiplicité il n'est nécessaire d'utiliser que une seule fiche comportant un intitulé complet (Fig.1); le cartouche des suivantes étant plus simple (Fig.2); il ne rapelle que les indications de niveau, de carré et de gisement. Ceci bien entendu dans le cas d'une seule collection, d'un même gisement.

BIBLIOGRAPHIE

LAPLACE G. - 1972 - La typologie analytique et structurale: base rationnelle d'étude des industries lithiques et osseuses. Colloques nationaux C.N.R.S., n.932. Banques des données archéologiques, Marseille, pp. 91-143.

MORELON S. - 1973 - Le gisement préhistorique de "La Touvière" commune d'Arbignieu (Ain). Doc.Labo.Géol.Fac.sciences Lyon , n.56, 165 pp., 61 figg., 5 tabl., 8 pl.

Fig.1

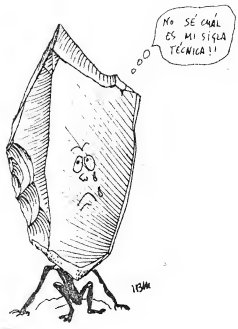
PAYS: FRANCE	DEPART: 01	COMM. Neuville S/Ain	GISEMENT: La Colombière
Fouilleurs Mayet et Pissot	Date 1915	Coll.	
Niveau D	Carré	Mayet et Pissot. 1915. L'abri sous roche préhist. de la Colombière près Poncin - Ain -	

OUTILS ret. compl.	Surélevée	Simple	Abrupte	Plate	Profonde	Marginale	directe	inverse	alterne	couvrante	envahissante	sommaire	scalarif.	lamellaire	dextre	sénestre	distale	proximale	médiane	divergente	convergente	concave	convexe	Ind.allong.	Ind.carenage	observations --- dimensions	n. coll.
G 1.2.		×			×		×																×			70-38-10	250[illegible]
—[R2.		×			×		×				(	)															
R2		×			×		×					×			×			×	×								
R1]		×				×	×					×			×		×		×				×				
+E1				×	×			×				(	)										×				
G 1.1.		×			×		×					(	)	×			×									40-15-0[illegible] cass prox.	250227
G 1.1.		×			×			×						×			×									65-20-15	250228
•T2.1			×		×		×						×					×				×					
G 1.1.		×			×		×										×									3[illegible]	250229
•Bc 1																											
[T2.2			×		×		×					(	)			×		×									
/D.1.1]		×			×		×								×			×									

Fig.2

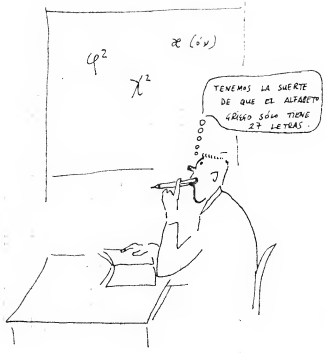
Niveau	Carré	Gisement																									
OUTILS ret. compl.	Surélevée	Simple	Abrupte	Plate	Profonde	Marginale	directe	inverse	alterne	couvrante	envahissante	sommaire	scalarif.	lamellaire	dextre	sénestre	distale	proximale	médiane	divergente	convergente	concave	convexe	Ind.allong.	Ind.carenage	observations --- dimensions	n. coll.

ABUDY, 1975

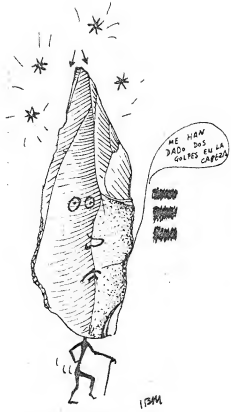


Les 4 Modes de la Retouche:

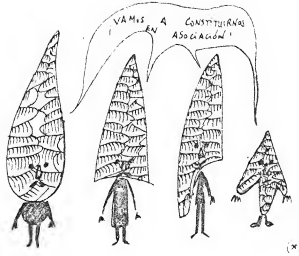
1) PLOP, PLOP, PLOP
2) TOC, TOC, TOC
3) CHOF, CHOF, CHOF.
4) TAT, TAT, TAT



ARUDY, 1975



Monsieur Bipan



F.E.T. (*)

(*) Foliacés sur Eclat Fronqué